



1. DATOS BÁSICOS DEL TFG:

Título: Razonamiento simbólico en modelos de lenguaje de gran tamaño: un caso de estudio en Matemáticas

Descripción general (resumen y metodología):

El presente trabajo tiene como objetivo evaluar la capacidad de los modelos de lenguaje de gran tamaño (LLMs, por sus siglas en inglés) para realizar razonamiento simbólico en áreas del Grado en Matemáticas. A partir de problemas representativos de asignaturas como Álgebra, Cálculo, Teoría de Conjuntos o Álgebra Lineal, se construirá un banco de pruebas con el que se evaluará la precisión, coherencia y solidez de las respuestas proporcionadas por modelos como GPT-4 o Qwen. Este trabajo se sitúa en la intersección entre la inteligencia artificial y las matemáticas, y busca no solo medir el rendimiento de estos modelos, sino también comprender los tipos de errores más frecuentes, las limitaciones actuales en el razonamiento simbólico automático y el potencial de estas herramientas.

La metodología que guiará el desarrollo del trabajo constará de los siguientes pasos:

1. Revisión bibliográfica sobre modelos de lenguaje de gran tamaño, matemáticas y razonamiento simbólico
2. Selección de un dominio de aplicación y problemas representativos
3. Definición de la metodología específica de la evaluación: "prompting", enfoque cualitativo y cuantitativo, etc.
4. Evaluación de modelos de lenguaje y análisis de errores
5. Redacción de la memoria del trabajo

Tipología: Estudio de casos, teóricos o prácticos, relacionados con la temática del Grado.

Objetivos planteados:

- Conocer los fundamentos teóricos de los modelos de lenguaje de gran tamaño (LLMs) y su aplicación al razonamiento simbólico.
- Elaborar un conjunto de problemas adecuados al nivel de matemáticas de grado.
- Evaluar el rendimiento de distintos LLMs en estos problemas, usando diferentes estrategias de interacción.
- Analizar y categorizar los errores cometidos por los modelos.
- Reflexionar sobre las implicaciones y las limitaciones del razonamiento simbólico en LLMs.

Bibliografía básica:

1. Simon Frieder, Luca Pinchetti, Alexis Chevalier, et al., "Mathematical Capabilities of ChatGPT," arXiv preprint arXiv:2301.13867, 2023.
2. Dan Hendrycks, Collin Burns, Saurav Kadavath, et al., "Measuring Mathematical Problem Solving With the MATH Dataset," arXiv preprint arXiv:1912.01412, 2021.
3. Qwen Team, Qwen2-Math: Large Language Models for Advanced Mathematical Reasoning, 2024. [Online]. Available: <https://qwenlm.github.io/blog/qwen2-math/>. Accessed: Apr. 9, 2025.

4. Qwen Team, Qwen2.5 Math Demo, 2024. [Online]. Available: <https://huggingface.co/spaces/Qwen/Qwen2.5-Math-Demo>. Accessed: Apr. 9, 2025.
5. Zhuosheng Zhang, Yuwei Liu, Jifan Xu, et al., "MathVista: Evaluating Mathematical Reasoning of Foundation Models in Visual Contexts," arXiv preprint arXiv:2212.10535, 2022.

Recomendaciones y orientaciones para el estudiante:

Plazas: 1

2. DATOS DEL TUTOR/A:

Nombre y apellidos: JUAN GÓMEZ ROMERO

Ámbito de conocimiento/Departamento: CIENCIAS DE LA COMPUTACIÓN E INTELIGENCIA ARTIFICIAL

Correo electrónico: jgomez@ugr.es

3. COTUTOR/A DE LA UGR (en su caso):

Nombre y apellidos:

Ámbito de conocimiento/Departamento:

Correo electrónico:

4. COTUTOR/A EXTERNO/A (en su caso):

Nombre y apellidos:

Correo electrónico:

Nombre de la empresa o institución:

Dirección postal:

Puesto del tutor en la empresa o institución:

Centro de convenio Externo:

5. DATOS DEL ESTUDIANTE:

Nombre y apellidos: LAURA GRAGERA PÉREZ

Correo electrónico: lauragragera@correo.ugr.es